stichting
mathematisch
centrum

MC

AFDELING MATHEMATISCHE BESLISKUNDE
(DEPARTMENT OF OPERATIONS RESEARCH)

BW 67/76 NOVEMBER

A. FEDERGRUEN

ON N-PERSON STOCHASTIC GAMES WITH DENUMERABLE STATE SPACE

2e boerhaavestraat 49 amsterdam

Printed at the Mathematical Centre, 49, 2e Boerhaavestraat, Amsterdam.

The Mathematical Centre, founded the 11-th of February 1946, is a non-profit institution aiming at the promotion of pure mathematics and its applications. It is sponsored by the Netherlands Government through the Netherlands Organization for the Advancement of Pure Research (Z.W.O), by the Municipality of Amsterdam, by the University of Amsterdam, by the Free University at Amsterdam, and by industries.

AMS(MOS) subject classification scheme (1970): 93E05

On N-person stochastic games with denumerable state space

by

A. Federgruen

ABSTRACT

This paper considers non-cooperative N-person stochastic games with a countable state space and compact metric action spaces. We concentrate upon the average return per unit time criterion for which the existence of an equilibrium policy is established under a number of recurrency conditions with respect to the transition probability matrices associated with the stationary policies. These results are obtained by establishing the existence of total discounted return equilibrium policies, for each discount-factor $\alpha \in [0,1)$ and by showing that under each one of the aforementioned recurrency conditions, average return equilibrium policies appear as limit policies of sequences of discounted return equilibrium policies, with discount factor tending to one.

Finally, we review and extend the results that are known for the case where both the state space and the action spaces are finite.

KEY WORDS & PHRASES: *non-cooperative stochastic games, denumerable state space, compact metric action spaces, equilibrium policies, average and total discounted return criterium.*

0. INTRODUCTION AND SUMMARY

A huge literature on Markov Decision Problems exists, in which a single decision maker controls the development of some system. However, in many stochastic control problems arising in various applications such as the modelling of economic markets, the description of biological systems etc. (cf. SOBEL [23]), the system is simultaneously controlled by more than one decisionmaker. As a consequence these problems have to be modelled using stochastic games.

This paper considers non-cooperative N-person stochastic games with a countable state space and compact metric action spaces. We concentrate upon the average return per unit time criterion for which both existence of an equilibrium policy and solutions to the optimality equation are established, under a number of recurrency conditions with respect to the transition probability matrices associated with the stationary policies.

These results are obtained by showing that the average return criterion arises as a (first) sensitive discount optimality criterion. More specifically we show that under each one of the aforementioned recurrency conditions, average return equilibrium policies appear as limit policies of sequences of total discounted return equilibrium policies with discount factor tending to one.

Accordingly, after giving some preliminaries and notation in section 1, we first establish in section 2 the existence of a total discounted return equilibrium policy for each discountfactor $\alpha \in [0,1)$ (an existing proof in [23] appears to be incorrect).

In section 3, existence of an average return equilibrium policy and a solution to the optimality equation are established, whereas in section 4, we review and extend the results that are known for the case where both the state space and the action spaces are finite.

1. PRELIMINARIES AND NOTATION

This paper treats an N-person noncooperative stochastic game specified by the objects S, $A^i(s)$, q and r.

S is a countable set, and for each i = 1,...,N and s ∈ S, $A^i(s)$ is a compact metric space where the set S denotes the state space of some system, and $A^i(s)$ denotes the set of actions, available to player i, in state s.

We define A as the union of all $A^i(s)$ (s∈S; i=1,...,N) and C as

$$C = X_{i=1}^{N} A. \tag{1.1}$$

q associates with each pair $(s,\underline{a}) \in S \times C$ a probability distribution $q_{s.}(\underline{a})$ on the elements of S; and r^i is a bounded real-valued function on $S \times C$, for all i = 1,...,N.

A stochastic game may be considered as a sequence $\gamma_1,\gamma_2,\ldots$ of non-cooperative games played by the N players, where s ∈ S indexes the set $\{\Gamma_s \mid s \in S\}$ from which γ_t (t=1,2,...) is drawn. Note that all the players' actions in $\gamma_t = s$ (t=1,2,..., s∈S) constitute a vector $\underline{a} = [a^1,\ldots,a^N] \in$ ∈ C(s) where

$$C(s) = X_{i=1}^{N} A^i(s), \qquad s \in S. \tag{1.2}$$

When $\gamma_t = s$ i.e. when the system is in state s and the vector $\underline{a} \in C(s)$ denotes all the players' actions in γ_t, then the one-step expected reward to player i, is given by $r^i(s;\underline{a})$ and the system moves to state t with probability $q_{st}(\underline{a})$.

For each s ∈ S, and i = 1,...,N let $F(A^i(s))$ denote the set of all signed measures on $\mathcal{B}_{A^i(s)}$, the Borel subsets of $A^i(s)$, endowed with the weak topology (cf. VARADARAJAN [27], p.16-17). The sets belonging to the base by which this topology is defined satisfy the Hausdorff postulates for neighbourhoods, and are in addition locally convex. As a consequence we obtain that $F(A^i(s))$ is a linear Hausdorff locally convex topological space.

Let $M(A^i(s))$ be the subspace of all <u>probability</u> measures on $\mathcal{B}_{A^i(s)}$, with the induced topology. It then follows from th. 3.4 in [27] that $M(A^i(s))$ can be metrized as a compact convex metric subspace of $F(A^i(s))$ since $A^i(s)$ is a compact metric space.

Next we define for each s ∈ S, $F(C(s)) = X_{i=1}^{N} F(A^i(s))$, and $M(C(s)) = X_{i=1}^{N} M(A^i(s))$, i = 1,...,N.

Note that F(C(s)) is again a linear Hausdorff locally convex topological space, and that M(C(s)) is again a compact convex metrizable subspace

of $F(C(s))$, $s \in S$. Finally, we observe that $M(C(s))$ can be identified as the space of all product probability measures on $\mathcal{B}_{C(s)}$, the product σ-field in $C(s)$. Moreover, for any sequence $\{\underline{\mu}_n\}_{n=1}^{\infty}$ with $\underline{\mu}_n \in M(C(s))$, $n = 1,2,\ldots$ it follows from th. 3.2 in BILLINGSLEY [3] that

$$(1.3) \qquad \int_{C(s)} r(\underline{a})\, d\,\mu_n(\underline{a}) \to \int_{C(s)} r(a)\, d\underline{\mu}(a), \qquad \text{as } n \to \infty$$

for all real-valued and continuous functions $r(\cdot)$ on $C(s)$ if and only if $\underline{\mu}_n \to \underline{\mu}$ (in the **product** topology).

We use the (abbreviated) notation $[\underline{\mu}^{-i},\nu]$ for the N-person randomized action $[\mu^1,\ldots,\mu^{i-1},\nu,\mu^{i+1},\ldots,\mu^N]$ that results from $\mu = [\mu^1,\ldots,\mu^N]$ when the i-th player changes from μ^i to ν, the other players continuing to use their respective actions in $\underline{\mu}$. Defining $r^i(s;\underline{\mu}) = E_{\underline{\mu}} r^i(s;\underline{a})$ and $q_{st}(\underline{\mu}) = E_{\underline{\mu}}\, q_{st}(\underline{a})$ for all $\underline{\mu} = [\mu^1,\ldots,\mu^N] \in M(C(s))$, $s \in S$, $i = 1,\ldots,N$, we obtain

$$(1.4) \qquad r^i(s;\underline{\mu}) = \int_{C(s)} r^i(s;\underline{a})\, d\mu(\underline{a}) =$$

$$= \int_{A^1(s)} \cdots \int_{A^N(s)} r^i(s;a^1,\ldots,a^N)\, d\mu^1(a^1)\ldots d\mu^N(a^N)$$

$$(1.5) \qquad q_{st}(\underline{\mu}) = \int_{C(s)} q_{st}(\underline{a})\, d\underline{\mu}(\underline{a}) =$$

$$= \int_{A^1(s)} \cdots \int_{A^N(s)} q_{st}(a^1,\ldots,a^N)\, d\mu^1(a^1)\ldots d\mu^N(a^N)$$

where the second equality in (1.4) and (1.5) follows from Fubini's theorem. Observe that $r^i(s;\underline{\mu})$ and $q_{st}(\underline{\mu})$ are both multilinear in μ, i.e. for all $\lambda \in [0,1]$:

$$(1.6) \qquad r^i(s;\mu^1,\ldots,\lambda\mu^j+(1-\lambda)\nu^j,\ldots,\mu^N) = \lambda r^i(s;\mu^1,\ldots,\mu^j,\ldots,\mu^N) +$$

$$+ (1-\lambda)\, r^i(s;\mu^1,\ldots,\nu^j,\ldots,\mu^N)$$

$$(1.7) \qquad q_{st}(\mu^1,\ldots,\lambda\mu^j+(1-\lambda)\nu^j,\ldots,\mu^N) = \lambda q_{st}(\mu^1,\ldots,\mu^j,\ldots,\mu^N) +$$

$$+ (1-\lambda)\, q_{st}(\mu^1,\ldots,\nu^j,\ldots,\mu^N).$$

Hereafter we assume that for each $s \in S$,

(1.8) $r^i(s;\underline{a})$ and $q_{st}(\underline{a})$ are continuous on $C(s)$, for all $i = 1,\ldots,N$ and $t \in S$.

Observe from (1.3) that (1.8) implies that for each $s \in S$, the one-step expected rewards and transition probabilities are continuous on the space of all *randomized* N players' actions $M(C(s))$ as well:

(1.9) $r^i(s;\underline{\mu})$ and $q_{st}(\underline{\mu})$ are continuous on $M(C(s))$ for all $i = 1,\ldots,N$ and $t \in S$.

Let $\Delta^i = X_{s\in S} M(A^i(s))$ be the set of all *decision rules for player* i, $(i=1,\ldots,N)$ i.e. of all functions δ^i mapping each state s into an action $\delta^i_s \in M(A^i(s))$. A policy for a player i is a sequence $\pi^i = (\delta^i(1),\delta^i(2),\ldots)$ of decision rules. Using policy π^i means that $\delta^i(n)$ is employed at time n; thus if the system is observed in state s at time n, then player i chooses action $\delta^i_s(n)$, the s-th component of $\delta^i(n)$. We write $\delta^{i(\infty)}$ for the *stationary policy* $(\delta^i,\delta^i,\ldots)$ for player i.

As a consequence we let Δ^i represent the class of all stationary policies for player i as well.

A stationary policy $\delta^{i(\infty)} \in \Delta^i$ is said to be *pure* if in each state $s \in S$ it prescribes a specific action in $A^i(s)$ with probability one.

Finally, the set of all policies for player i is denoted by Π^i, and $\underline{\Pi} = X_{i=1}^N \Pi^i$ represents the class of all N players' policies, with $\underline{\Delta} = X_{i=1}^N \Delta^i$ the subset of the *stationary* N players' policies.

We associate with each stationary policy $\underline{\delta}^{(\infty)} \in \underline{\Delta}$, the transition probability matrix $P(\underline{\delta})$, i.e.

$$P(\underline{\delta})_{st} = q_{st}(\underline{\delta}(s))$$

with the n-th power $P^n(\underline{\delta})$ indicating the matrix of n-step transition probabilities, i.e. $P^n(\underline{\delta}) = P(\underline{\delta})\, P^{n-1}(\underline{\delta})$, $n \geq 2$.

For any policy $\underline{\pi} = [\pi^1,\ldots,\pi^N] \in \underline{\Pi}$ we define $V^i_\alpha(\underline{\pi},s)$ and $g^i(\underline{\pi},s)$ as

the total expected α-discounted return, and the long-run average return per unit time to player i, when the initial state is s:

$$(1.10)\qquad V_\alpha^i(\underline{\pi};s) = E_{\underline{\pi}}\{\sum_{k=0}^{\infty}\alpha^k r^i(s_k;\underline{a}_k) \mid s_0=s\};\ i = 1,\ldots,N;\ s \in S;\ 0 \le \alpha < 1$$

$$(1.11)\qquad g^i(\underline{\pi};s) = \lim_{t\to\infty}\sup \frac{1}{t+1} E_{\underline{\pi}}\{\sum_{k=0}^{t} r^i(s_k;\underline{a}_k) \mid s_0=s\};\ i = 1,\ldots,N;\ s \in S$$

where $E_{\underline{\pi}}$ indicates the expectation given the players' common policy $\underline{\pi} \in \underline{\Pi}$ is used and where $\{s_k;k=0,1,2,\ldots\}$ and $\{\underline{a}_k;k=0,1,\ldots\}$ denote the stochastic processes of the states and actions that result from policy $\underline{\pi}$.

An N-tuple of policies $\underline{\pi}^* = [\pi^{*1},\ldots,\pi^{*N}] \in \underline{\Pi}$ is said to be an *α-discounted equilibrium point of policies* (α-DEP) if, *simultaneously for every initial state of the system* s,

$$(1.12)\qquad V_\alpha^i(\underline{\pi}^*;s) \ge V_\alpha^i(\underline{\pi};s) \text{ for all } i = 1,\ldots,N \text{ and } \underline{\pi} \in \Pi^{-i}(\underline{\pi}^*):$$

where

$$(1.13)\qquad \Pi^{-i}(\underline{\pi}^*) = \{\underline{\pi} = [\pi^1,\ldots,\pi^N] \in \underline{\Pi} \mid \pi^j = \pi^{*j},\ j \ne i\}.$$

Similarly we define $\underline{\pi}^*$ as an *average return equilibrium point of policies (AEP)*, if *simultaneously for every initial state* s,

$$(1.14)\qquad g^i(\underline{\pi}^*;s) \ge g^i(\underline{\pi};s) \text{ for all } i = 1,\ldots,N \text{ and } \underline{\pi} \in \Pi^{-i}(\underline{\pi}^*).$$

Hence, whenever the players choose an α-DEP (AEP) $\underline{\pi}^*$, none of them, whatever the initial state of the system, can increase his own total expected α-discounted return (expected average return per unit time) by changing to some other policy $\pi^i \ne \pi^{*i} \in \Pi^i$, the other players continuing to use their respective policies in $\underline{\pi}^*$.

Note that we do not consider *history-dependent* policies, i.e. policies which prescribe for each time t, a randomized action in dependence on the entire history $H_t = (s_0;\underline{a}_0;a_1,\ldots,s_{t-1},\underline{a}_{t-1},s_t)$ of the system up to time t, rather than in dependence on the current state s_t alone. The justification for our confining ourselves to the class $\underline{\Pi}$ is provided by [13], who showed

as an adaptation of the corresponding result in DERMAN & STRAUCH [7] that whenever a policy $\underline{\pi}^*$ is an α-DEP or AEP within $\underline{\Pi}$, it is an equilibrium policy within the broader class of history-dependent policies as well.

We conclude this section by observing that if the sets $A^i(s)$ $(i=1,\ldots,N;\ s\in S)$ are convex compact subsets of some linear metric space themselves, such that for all $i = 1,\ldots,N$ $r^i(s;\underline{a})$ is linear or even concave in the i-th component of $\underline{a}$ (cf. (1.6) and (1.7)) then the existence of a *pure* instead of a randomized stationary α-DEP or AEP is guaranteed under the same conditions as follows from an examination of the analysis below.

2. EXISTENCE OF STATIONARY α-DEP'S

In this section we prove the existence of a stationary α-DEP for each $\alpha \in [0,1)$*).

For each policy $\underline{\delta}^{(\infty)} \in \underline{\Delta}$, the total expected α-discounted return to player i, when starting in state $s \in S$, is denoted by

$$(2.1) \qquad V_\alpha^i(\underline{\delta}^{(\infty)};s) = \sum_{n=0}^{\infty} \alpha^n \sum_{t\in S} P^n(\underline{\delta})_{st}\, r^i(t;\underline{\delta}(t)).$$

The following lemma proves that $V_\alpha^i(\underline{\delta}^{(\infty)};s)$ is a continuous function on $\underline{\Delta}$ for all $i = 1,\ldots,N$, $s \in S$ and $\alpha \in [0,1)$:

LEMMA 2.1. *Fix* $s \in S$, $1 \le i \le N$ *and* $\alpha \in [0,1)$. *Then* $V_\alpha^i(\underline{\delta}^{(\infty)};s)$ *is continuous on* $\underline{\Delta}$.

PROOF. We first observe that since $\underline{\Delta}$ is metrizable, it suffices to show that $\lim_{n\to\infty} V_\alpha^i(\underline{\delta}_n^{(\infty)};s) = V_\alpha^i(\underline{\delta}^{(\infty)};s)$ whenever $\{\underline{\delta}_n\}_{n=1}^{\infty} \to \underline{\delta}$, with $\underline{\delta}_n \in \underline{\Delta}$. Fix a sequence $\{\underline{\delta}_n\}_{n=1}^{\infty}$ with $\lim_{n\to\infty} \underline{\delta}_n = \underline{\delta}$ and note that $\underline{\delta} \in \underline{\Delta}$, in view of the compactness of $\underline{\Delta}$. Let M be such that

$$(2.2) \qquad |r^i(s;\underline{a})| \le M \quad \text{for all } s \in S, \text{ and } \underline{a} \in C(s).$$

It is then easily verified that

$$(2.3) \qquad |V_\alpha^i(\underline{\eta}^{(\infty)};s)| \le M/(1-\alpha) \quad \text{for all } \underline{\eta}^{(\infty)} \in \underline{\Delta} \text{ and } s \in S.$$

---

*) Shortly after completing this paper, I became aware from a recent bibliography of a report by IDZIK [14] in which similar existence results for α-DEPs seem to be obtained.

Next, observe by complete induction that as a consequence of (1.8) $P^k(\underline{\delta})_{st}$ is continuous on $\underline{\Delta}$ for all $s,t \in S$ and $k = 1,2,\ldots$. This, in turn, implies using proposition 18 on p.232 in ROYDEN [18] that for each $\ell = 0,1,\ldots$

$$(2.4) \qquad \lim_{n\to\infty} \sum_t P^{\ell}(\underline{\delta}_n)_{st}\, r^i(t;\underline{\delta}_n(t)) = \sum_t P^{\ell}(\underline{\delta})_{st}\, r^i(t;\underline{\delta}(t)).$$

Finally, pick $\varepsilon > 0$ and choose K such that $\alpha^K \le \varepsilon(1-\alpha)/4M$. Let $H^k_{\underline{\eta}}(s) = \sum_{\ell=0}^{k-1} \alpha^{\ell} \sum_{t\in S} P^{\ell}(\underline{\eta})_{st}\, r^i(t;\underline{\eta}(t))$ for all $k = 1,2,\ldots$ and $\underline{\eta} \in \underline{\Delta}$. Observe that for each $\underline{\eta} \in \underline{\Delta}$:

$$(2.5) \qquad V^i_{\alpha}(\underline{\eta}^{(\infty)};s) = H^K_{\underline{\eta}}(s) + \alpha^K \sum_{t\in S} P^K(\underline{\eta})_{st}\, V^i_{\alpha}(\underline{\eta}^{(\infty)};t).$$

In view of (2.4) there exists an integer N_0 such that $|H^K_{\underline{\delta}_n}(s) - H^K_{\underline{\delta}}(s)| \le \varepsilon/2$, for all $n \ge N_0$. We thus obtain that for all $n \ge N_0$:

$$|V^i_{\alpha}(\underline{\delta}_n^{(\infty)};s) - V^i_{\alpha}(\underline{\delta}^{(\infty)};s)| \le |H^K_{\underline{\delta}_n}(s) - H^K_{\underline{\delta}}(s)| +$$

$$+ \alpha^K \Big| \sum_{t\in S} P^K(\underline{\delta}_n)_{st}\, V^i_{\alpha}(\underline{\delta}_n^{(\infty)};t) - \sum_{t\in S} P^K(\underline{\delta})_{st}\, V^i_{\alpha}(\underline{\delta}^{(\infty)};t)\Big| \le$$

$$\varepsilon/2 + \frac{\varepsilon(1-\alpha)}{4M}\,\frac{2M}{(1-\varepsilon)} = \varepsilon. \qquad \square$$

We now turn to the existence of an α-DEP.
For a compact, metric state space and under somewhat stronger continuity assumptions with respect to the one-step expected rewards, and transaction probability functions, the existence of an α-DEP was first proved by SOBEL [23]. Unfortunately there seem to be a number of serious errors which invalidate the approach. Although with a considerable amount of additional work, the proof in [23] can be rectified for the case of a denumerable state space we prefer to give a different proof.

Our approach uses an extension of the Kakutani fixed-point theorem which was obtained independently by GLICKSBERG [11] and FAN [9]. First, for each compact set U let 2^U denote the class of all (non-empty) closed subsets of U. A point to set mapping $\Phi: U \to 2^U$ (with U satisfying the first countability axiom) is said to be upper semi-continuous, if for each

sequence $\{x_n\}_{n=1}^{\infty}$, $x_n \in U$:

$$(2.6) \qquad \{x_n\}_{n=1}^{\infty} \to x,\ y_n \in \Phi(x_n),\ \{y_n\}_{n=1}^{\infty} \to y \Rightarrow y \in \Phi(x).$$

LEMMA 2.2. *Given an upper semi-continuous point to convex set mapping* $\Phi: U \to U$ *of a convex compact subset* U *of a linear Hausdorff locally convex topological space into itself, there exists a point* $x \in \Phi(x)$. □

Observe from the analysis in section 1, that $X_{s\in S} F(C(s))$, the space of all functions f mapping each state s into a N-tuple of (finite, signed) measures $f_s \in F(C(s))$, endowed with the product topology is again a linear Hausdorff locally convex topological space, with $\underline{\Delta}$, the countable topological product of the spaces M(C(s)) (s∈S), a metrizable subspace which is in addition convex and compact, as a consequence of Tychonoff's theorem. The fixed point theorem in lemma 2.2 will be applied by constructing a point to set mapping on $\underline{\Delta}$, as a subspace of $X_{s\in S} F(C(s))$.

We finally need the following lemma, the proof of which follows from th. 6-f in BLACKWELL [5]:

LEMMA 2.3. *Fix* $0 \le \alpha < 1$. *A stationary policy* $\underline{\delta}^{(\infty)} = [\delta^{1(\infty)},\ldots,\delta^{N(\infty)}]$ *is an* α-DEP, *iff* $V_\alpha^i(\underline{\delta}^{(\infty)};s)$ *satisfies the optimality equation:*

$$(2.7) \qquad V_\alpha^i(\underline{\delta}^{(\infty)};s) = \max_{\mu\in M(A^i(s))} \{r^i(s;[\delta^{-i}(s),\mu]) + \alpha \sum_{t\in S} q_{st}([\delta^{-i}(s),\mu])\, V_\alpha^i(\underline{\delta}^{(\infty)};t)\}$$

for all $s \in S$, $i = 1,\ldots,N$.

THEOREM 1. *There exists a stationary* α-DEP *for each* $\alpha \in [0,1)$.

PROOF. We first observe that for each $\underline{\delta} \in \underline{\Delta}$ and $i = 1,\ldots,N$ there exists, as a result of (1-8) an $\eta \in \Delta^i$ such that for all $s \in S$:

$$(2.8) \qquad r^i(s;[\delta^{-i}(s),\eta(s)]) + \alpha \sum_{t\in S} q_{st}([\delta^{-i}(s),\eta(s)])\, V_\alpha^i(\underline{\delta}^{(\infty)};t) = \max_{\mu\in M(A^i(s))} \{r^i(s;[\delta^{-i}(s),\mu]) + \alpha \sum_{t\in S} q_{st}([\delta^{-i}(s),\mu])\, V_\alpha^i(\underline{\delta}^{(\infty)};t)\}.$$

For any $i = 1,\ldots,N$ and $\underline{\delta} \in \underline{\Delta}$, let $\Phi^i(\underline{\delta})$ denote the set of all $\eta \in \Delta^i$ that satisfy (2.9) for all $s \in S$, and define the point-to-convex set mapping

$$\Phi: \underline{\Delta} \to 2^{\underline{\Delta}}: \underline{\delta} \to \Phi(\underline{\delta}) = X_{i=1}^{N} \Phi^i(\underline{\delta}).$$

We next show the upper-semi-continuity of this point-to-set mapping. Fix $\{\underline{\delta}_n\}_{n=1}^{\infty}$, $\{\underline{\eta}_n\}_{n=1}^{\infty}$ with (1) $\underline{\delta}_n$, $\underline{\eta}_n \in \underline{\Delta}$, (2) $\lim_{n\to\infty} \underline{\delta}_n = \underline{\delta}$; $\lim_{n\to\infty} \underline{\eta}_n = \underline{\eta}$ and (3) $\underline{\eta}_n \in \Phi(\underline{\delta}_n)$.
Substitute $\underline{\delta}_n$ for $\underline{\delta}$ and η_n^i for η in (2.8) and let n tend to infinity. It then follows that η^i satisfies (2.8) for $\underline{\delta}$, and this for all $i = 1,\ldots,N$ and $s \in S$, as a consequence of (1.8), lemma 2.2, the boundedness of $V_\alpha^i(\underline{\delta}_n^{(\infty)};s)$ and proposition 18 on p.232 in ROYDEN [18].

As a consequence of the upper-semi-continuity of Φ, and the fact that Φ is a point-to-convex set mapping of a convex compact subset $\underline{\Delta}$ of the linear Hausdorff locally convex topological space $X_{s\in S} F(C(s))$ into itself, it follows from lemma 2.2 that there exists a $\underline{\delta}^* \in \underline{\Delta}$ such that $\delta^* \in \Phi(\underline{\delta}^*)$ which implies (2.7) and hence proves the theorem (cf. lemma 2.3). □

3. THE EXISTENCE OF AVERAGE RETURN EQUILIBRIUM POLICIES (AEP'S)

We first introduce the following notation:
For any $\underline{\delta} \in \underline{\Delta}$ we define the matrix $P^*(\underline{\delta})$ as the Cesaro limit of the sequence $\{P^n(\underline{\delta})\}_{n=1}^{\infty}$. For each pair of states $s,t \in S$, we denote by $m_{\underline{\delta}}(s,t)$ the mean first passage time, i.e. the expected number of transitions needed to get from state s to state t, under policy $\underline{\delta}^{(\infty)} \in \underline{\Delta}$. Let $R(\underline{\delta})$ denote the set of recurrent states under $P(\underline{\delta})$ and recall that a state $t \in R(\underline{\delta})$ is positive recurrent if and only if $P^*(\underline{\delta})_{tt} > 0$. For each state $s \in R(\underline{\delta})$, let $\lambda(\underline{\delta}) \geq 1$ denote the period of the *subchain* (closed, irreducible set of states) to which state s belongs, and if $R(\underline{\delta}) \neq \emptyset$ let $\lambda(\underline{\delta})$ be the least common multiple of the periods of the different subchains (which may be infinite). We next define for each pair of states $s,t \in S$; each policy $\underline{\delta} \in \underline{\Delta}$ and $r = 1,\ldots,\lambda(\underline{\delta})$:

$f_{st}^*(r;\underline{\delta})$ as the probability that the system under $P(\underline{\delta})$ will ever reach state t in the course of the transitions r; $\lambda(\underline{\delta}) + r$; $2\lambda(\underline{\delta}) + r,\ldots$ when

starting in state s. Note from th. 4 on p.31 in CHUNG [6] that for each $\underline{\delta} \in \underline{\Delta}$ and $s,t \in S$:

$$(3.1) \qquad \lim_{N\to\infty} P^{n\lambda(\underline{\delta})+r}(\underline{\delta})_{st} = \begin{cases} \lambda(\underline{\delta})_t \cdot f^*_{st}(r;\underline{\delta})\ P^*(\underline{\delta})_{tt} & \text{if } t \in R(\underline{\delta}) \\ 0 & \text{otherwise.} \end{cases}$$

If $P(\underline{\delta})$ has a single subchain the states of which are recurrent, let $\{C^m(\underline{\delta}) \mid m = 1,\ldots,\lambda(\underline{\delta})\}$ denote the set of cyclically moving subsets (*c.m.s.*) of this recurrent class which are numbered in such a way that for any $m = 1,\ldots,\lambda(\underline{\delta})$:

$$(3.2) \qquad s \in C^m(\underline{\delta}) \Rightarrow P(\underline{\delta})_{st} > 0 \quad \text{only if } t \in C^{m+1}(\underline{\delta})$$

with the convention that the superscript m in $C^m(\underline{\delta})$ is taken modulo $\lambda(\underline{\delta})$. Note that the sets $\{C^m(\underline{\delta}) \mid m = 1,\ldots,\lambda(\underline{\delta})\}$ are the subchains of the matrix $P^{\lambda(\underline{\delta})}(\underline{\delta})$. Finally let $f^*(s;r,C^m(\underline{\delta}))$ with $s \in S$, $m,r = 0,\ldots,\lambda(\underline{\delta}) - 1$ indicate the probability that the system will eventually be absorbed in $C^m(\underline{\delta})$ in the course of the transitions $r;\ \lambda(\underline{\delta}) + r;\ e\lambda(\underline{\delta}) + r;\ldots$ when starting in state s. We recall from th. 3 on p.31 in [6] that for all $\underline{\delta} \in \underline{\Delta}$ and $s \in S$:

$$(3.3) \qquad f^*_{st}(r;\underline{\delta}) = f^*(s;r;C^m(\underline{\delta}) \text{ for all } t \in C^m(\underline{\delta}) \text{ and } m = 1,\ldots,\lambda(\underline{\delta}).$$

Finally if the single subchain of $P(\underline{\delta})$ is a *positive recurrent class* there exists a unique stationary probability distribution $\pi(\underline{\delta})$ such that $P^*(\underline{\delta})_{st} = \pi(\underline{\delta})_t$ for all $s,t \in S$.

Next we introduce a number of recurrency conditions, each of which will be shown to guarantee the existence of an AEP.

A1. For each $\underline{\delta} \in \underline{\Delta}$, $P(\underline{\delta})$ has a single subchain. In addition there exist integers, $\nu,d \geq 1$, a number $\rho > 0$ and for each $\underline{\delta} \in \underline{\Delta}$ a nonnegative matrix $Q(\underline{\delta})$ such that for each subset $A \subseteq S$:

$$(3.4) \qquad \Big| \sum_{t\in A} \{P^{nd}(\underline{\delta})_{st} - Q(\underline{\delta})_{st}\} \Big| \leq (1-\rho)^{[n/\nu]}; \qquad s \in S$$

where $[x]$ denotes the largest integer less than or equal to x.

A2. There exists a number R such that for each player $i = 1,\ldots,N$ and for any combination of stationary policies $\{\delta^1,\ldots,\delta^{i-1},\delta^{i+1},\ldots,\delta^N\}$ of the other players, there is a policy $\delta^i \in \Delta^i$ for player i, for which the mean first passage time $m_{\underline{\delta}}(s,t)$ from any state s to any state t, under policy $\underline{\delta} = [\delta^1,\ldots,\delta^N]$ is bounded by R, i.e. for each $\{\delta^1,\ldots,\delta^{i-1},\delta^{i+1},\ldots,\delta^N\}$ with $\delta^j \in \Delta^j$ for all $j \neq i$, there exists a $\delta^i \in \Delta^i$ such that

$$(3.5) \qquad m_{\underline{\delta}}(s,t) \leq R \quad \text{for all } s,t \in S \text{ where } \underline{\delta} = [\delta^1,\ldots,\delta^N].$$

We first exhibit a number of properties that follow from assumption A1:

LEMMA 3.1. *Assume* A1 *holds. Then*

(a) d *is a multiple of the period* $\lambda(\underline{\delta})$ *for each* $\underline{\delta} \in \underline{\Delta}$.

(b) *For each* $\underline{\delta} \in \underline{\Delta}$ *the unique subchain of* $P(\underline{\delta})$ *is a positive recurrent class.*

(c) *For each* $\underline{\delta} \in \underline{\Delta}$ *and* $r = 0,\ldots,d-1$ *there exist nonnegative matrices* $Q^{(r)}(\underline{\delta})$ *(with* $Q^{(0)}(\underline{\delta}) = Q(\underline{\delta})$*) such that for each subset* $A \subseteq S$*:*

$$(3.6) \qquad \Big|\sum_{t\in A}\{P^{nd+r}(\underline{\delta})_{st} - Q^{(r)}(\underline{\delta})_{st}\}\Big| \leq (1-\rho)^{[n/\nu]}; \qquad s \in S$$

where

$$(3.7) \qquad Q^{(r)}(\underline{\delta})_{st} = \begin{cases} 0 & \text{for all } t \notin R(\underline{\delta}) \\ \lambda(\underline{\delta}) f^*(s;r;C^m(\underline{\delta}))\pi(\underline{\delta})_t & \text{for all } t \in C^m(\underline{\delta}),\ m = 1,\ldots,d. \end{cases}$$

(d) *For each* $\underline{\delta} \in \underline{\Delta}$ *and* $r = 1,\ldots,\lambda(\underline{\delta})$*:*

$$\sum_{m=1}^{\lambda(\underline{\delta})} f^*(s;r;C^m(\underline{\delta})) = 1 \qquad \text{for all } s \in S$$

and

$$\sum_{t\in R(\underline{\delta})} P^n(\underline{\delta})_{st} \geq 1 - (1-\rho)^{[n/\nu.d]} \quad \text{for all } s \in S \text{ and } m = 1,2,\ldots .$$

(e) $P^*(\underline{\delta}) = \lim_{n\to\infty} 1/\lambda(\underline{\delta}) \sum_{r=1}^{\lambda(\underline{\delta})} P^{n\lambda(\underline{\delta})+r}(\underline{\delta})$ *and* $P^*(\underline{\delta})$ *depends continuously on* $\underline{\delta} \in \underline{\Delta}$*, i.e. for all* $s,t \in S$ $\lim_{\ell\to\infty} P^*(\underline{\delta}_\ell)_{st} = P^*(\underline{\delta})_{st}$ *whenever* $\{\underline{\delta}_\ell\} \to \{\underline{\delta}\}$.

PROOF.

(a) follows immediately from (3.1) and the fact that for each pair $s,t \in S$: $\lim_{n\to\infty} P^{nd}(\underline{\delta})_{st}$ exists.

(b) By choosing $A = \{t\}$ in (3.4) we obtain that $\lim_{n\to\infty} P^{nd}(\underline{\delta})_{st} = Q(\underline{\delta})_{st}$ for all $s,t \in S$; and by choosing $A = S$ in (3.4) it follows that $\sum_{t\in S} Q(\underline{\delta})_{st} = 1$ for all $s \in S$.
Assume to the contrary that the unique subchain of $P(\underline{\delta})$ contains transient or null-recurrent states; we then obtain in both cases, in view of part (a) and (3.1):

$$1 = \sum_{t\in S} Q(\underline{\delta})_{st} = \sum_{t\in S} \lim_{n\to\infty} P^{nd}(\underline{\delta})_{st} = 0$$

thus proving part (b) by contradiction.

(c) For $r = 0$ the assertion follows from the fact that $R(\underline{\delta})$ is a single positive recurrent class (cf. part (b)) and the combination of (3.1) and (3.3) as well as the fact that $Q(\underline{\delta})_{st} = \lim_{n\to\infty} P^{nd}(\underline{\delta})_{st}$ for all $s,t \in S$. Next, note that for $r = 1,\ldots,d$ and any subset $A \subseteq S$:

$$\Big|\sum_{t\in A} P^{nd+r}(\underline{\delta})_{st} - \sum_{t\in A}\{\sum_{u\in S} P^{r}(\underline{\delta})_{su}.Q(\underline{\delta})_{ut}\}\Big| =$$

$$\Big|\sum_{u\in S} P^{r}(\underline{\delta})_{su} \sum_{t\in A}\{P^{nd}(\underline{\delta})_{ut} - Q(\underline{\delta})_{ut}\}\Big| \leq$$

$$\leq \sum_{u\in S} P^{r}(\underline{\delta})_{su}(1-\rho)^{[n/\nu]} = (1-\rho)^{[n/\nu]}$$

thus showing the existence of nonnegative matrices $Q^{(r)}(\delta)$ which satisfy (3.6). The explicit expressions in (3.7) then follow again from the fact that $R(\underline{\delta})$ is a single positive recurrent class, and the combination of (3.1) and (3.3).

(d) Note that $\sum_{t\in C^m(\underline{\delta})} \pi(\underline{\delta})_t = 1/\lambda(\underline{\delta})$ for $m = 1,\ldots,\lambda(\underline{\delta})$ and use (3.7) to conclude that for each $s \in S$:

$$1 = \sum_{t\in S} Q^{(r)}(\underline{\delta})_{st} = \sum_{m=1}^{\lambda(\underline{\delta})} \lambda(\underline{\delta})\, f^{*}(s;r;C^{m}(\underline{\delta})) \sum_{t\in C^m(\underline{\delta})} \pi(\underline{\delta})_t =$$

$$= \sum_{m=1}^{\lambda(\underline{\delta})} \frac{\lambda(\underline{\delta})}{\lambda(\underline{\delta})} .f^*(s;r;C^m(\underline{\delta}) = \sum_{m=1}^{\lambda(\underline{\delta})} f^*(s;r;C^m(\underline{\delta}).$$

The second assertion follows from (3.6) with the choice $A = R(\underline{\delta})$ and the observation that $\sum_{t\in R(\underline{\delta})} Q^{(r)}(\underline{\delta})_{st} = 1$ (cf. (3.7)).

(e) Since $P^*(\underline{\delta}) = 1/\lambda(\underline{\delta}) \sum_{r=1}^{\lambda(\underline{\delta})} Q^{(r)}(\underline{\delta})$ it suffices to show the continuity of each of the matrix functions $Q^{(r)}(\underline{\delta})$ (r=1,...,d) on $\underline{\Delta}$. Fix $s,t \in S$ and observe in view of (3.6) that we can fix n sufficiently large such that uniformly for all $\underline{\delta} \in \underline{\Delta}$, $P^{nd+r}(\underline{\delta})_{st}$ comes arbitrarily close to $Q^{(r)}(\underline{\delta})_{st}$. Part (d) then follows from the continuity of $P^{nd+r}(\underline{\delta})_{st}$ on $\underline{\Delta}$, for all n = 1,2,... . □

Condition A1 is of course awkward to check; however there are a number of easily checkable and widely fitting recurrency conditions which imply A1, such as:

A1.1: There exists a finite set K, a positive integer n, and a positive real number c such that $\sum_{t\in K} P^n(\underline{\delta}) \geq c$ for all $s \in S$ and $\underline{\delta} \in \underline{\Delta}$. In addition, for each $\underline{\delta} \in \underline{\Delta}$, $P(\underline{\delta})$ has exactly one subchain.

A1.2: There is an integer $\nu \geq 1$ and a number $\rho > 0$ such that for each pair of states (s_1,s_2) and for each $\underline{\delta} \in \underline{\Delta}$:

$$\sum_{t=1}^{\infty} \min\{P^{\nu}(\underline{\delta})_{s_1t}, P^{\nu}(\underline{\delta})_{s_2t}\} \geq \rho. \tag{3.8}$$

A1.3: There exists a state s^* such that for each policy $\underline{\delta} \in \underline{\Delta}$, the mean first passage time $m_{\underline{\delta}}(s,s^*)$ is finite and uniformly bounded in $s \in S$, and $\underline{\delta} \in \underline{\Delta}$.

Both the first condition in A1.1 as the assumption A1.2 are generalizations of the Doeblin condition (cf. e.g. DOOB [8], p.197) to a collection of Markov chains; the former was introduced by HORDIJK [12] as the *simultaneous Doeblin condition*, and the latter is an adaptation of a condition introduced in TIJMS [26]. We note that A1.2 with $\nu = 1$ is equivalent to the condition that there is a number $\rho > 0$, such that for each four elements $(s_1,s_2,\underline{a}_1,\underline{a}_2)$ with $s_1 \neq s_2$ and $\underline{a}_1 \in C(s_1)$, $\underline{a}_2 \in C(s_2)$:

$$\sum_{t=1}^{\infty} \min\{q_{s_1t}(\underline{a}_1), q_{s_2t}(\underline{a}_2)\} \geq \rho. \tag{3.9}$$

For, fix $s_1, s_2 \in S$ and $\underline{\mu}_1 \in M(C(s_1))$, $\underline{\mu}_2 \in M(C(s_2))$ and observe that, as a consequence of (3.9):

$$(3.10) \qquad \rho \leq E_{\underline{\mu}_1,\underline{\mu}_2}\left[\sum_{t=1}^{\infty} \min\{q_{s_1 t}(\underline{a}_1),\ q_{s_2 t}(\underline{a}_2)\}\right] =$$

$$= \sum_{t=1}^{\infty} E_{\underline{\mu}_1,\underline{\mu}_2} \min\{q_{s_1 t}(\underline{a}_1), q_{s_2 t}(\underline{a}_2)\} \leq \sum_{t=1}^{\infty} \min\{q_{s_1 t}(\underline{\mu}_1), q_{s_2 t}(\underline{\mu}_2)\},$$

where the interchange of expectation and summation is justified by the non-negativity of $\min\{q_{s_1 t}(\underline{a}_1), q_{s_2 t}(\underline{a}_2)\}$, and where the inequality part follows from Jensen's inequality and the concaveness of $\min(.,.)$ on R^2. Note finally that (3.10) coincides with the special case of (3.8) where $\nu = 1$. In Markov Chain terminology, the condition (3.9) is known as the assumption that for each stationary and *pure* policy $\underline{\delta}^{(\infty)}$, the associated tpm $P(\underline{\delta})$ is scrambling (cf. [1]) and has an ergodic coefficient of at least ρ. Assumption A3.2 is an adaptation of a condition introduced in ROSS[17].

Note that both under A1.2 and A1.3 the tpm $P(\underline{\delta})$ of each $\underline{\delta} \in \underline{\Delta}$ has a single subchain, the states of which are positive recurrent. In a forthcoming paper we will show that assumption A1.1 implies A1 and the fact that A1.2 is a special case of A1 follows along lines with the proof of theorem 1 in ANTHONISSE & TIJMS [1].

Moreover in this same paper it will be shown that under the assumption that state s^* can be reached from any state s under any policy $\underline{\delta}^{(\infty)} \in \underline{\Delta}$, conditions A1, A1.1 and A1.3 are equivalent; this of course implies in particular that A1.3 is a special case of A1 as well.

We finally note (without proof) that under assumption A1, each one of the policies $\underline{\delta}^{(\infty)} \in \underline{\Delta}$ satisfies the Doeblin condition.

For each α $(0 \leq \alpha < 1)$ we choose a specific α-DEP $\underline{\delta}_\alpha \in \underline{\Delta}$. Next we fix any state s^* and define:

$$(3.11) \qquad v_\alpha^i(s) = V_\alpha^i(\underline{\delta}^{(\infty)}; s) - V_\alpha^i(\underline{\delta}_\alpha^{(\infty)}; s^*), \text{ for all } s \in S \text{ and } i = 1,\ldots,N.$$

LEMMA 3.2. *Both under assumption* A1 *and* A2, *the family of functions* $\{v_\alpha^i(\cdot);\ 0 \leq \alpha < 1\}$ *is uniformly bounded.*

PROOF. Under condition A2, the uniform boundedness of $\{v_\alpha^i(\cdot),\ i=1,\ldots,N\}$ follows from the proof of th.12.8 in HORDIJK [12]. To prove the lemma under condition A1, we show subsequently:

(a) $|V_\alpha^i(\underline{\delta}_\alpha^{(\infty)};s) - V_\alpha^i(\underline{\delta}_\alpha^{(\infty)};s')| \leq 4M\nu d/\rho$ for all $s,s' \in C^m(\underline{\delta}_\alpha)$, $m = 1,\ldots,\lambda(\underline{\delta}_\alpha)$; $\alpha \in [0,1)$.

(b) $|V_\alpha^i(\underline{\delta}_\alpha^{(\infty)};s) - V_\alpha^i(\underline{\delta}_\alpha^{(\infty)};s')| \leq (4\nu+2\rho)Md/\rho$ for all $s,s' \in R(\underline{\delta}_\alpha)$.

(c) $|V_\alpha^i(\underline{\delta}_\alpha^{(\infty)};s) - V_\alpha^i(\underline{\delta}_\alpha^{(\infty)};s')| \leq (6\nu+2\rho)Md/\rho$ for all $s \in S\backslash R(\underline{\delta}_\alpha)$ and $s' \in R(\underline{\delta}_\alpha)$.

Note that the assertion follows easily from the combination of (a), (b) and (c) for any choice of $s^* \in S$.
To prove (a), fix $\alpha \in [0,1)$, $1 \leq i \leq N$, $1 \leq r \leq \lambda(\underline{\delta}_\alpha)$ and $s,s' \in C^r(\underline{\delta}_\alpha)$. Let $\lambda = \lambda(\underline{\delta}_\alpha)$, and for any scalar a, let $a^+ = \max(a,0)$ and $a^- = \max(-a,0)$. Observe that for each $n = 0,1,\ldots$ and $m = 1,\ldots,\lambda$ there exist two sets A^+ and A^- with $A^+,A^- \subseteq S$ such that for each $s \in S$:

$$(3.12)\qquad \sum_{t\in S} \{P^{nd+m}(\underline{\delta}_\alpha)_{st} - Q^{(m)}(\underline{\delta}_\alpha)_{st}\}^\pm =$$

$$= \sum_{t\in A^\pm} \pm\{P^{nd+m}(\underline{\delta}_\alpha)_{st} - Q^{(m)}(\underline{\delta}_\alpha)_{st}\} \leq (1-\rho)^{[n/\nu]}$$

where the inequality follows from (3.6) with the choice $A = A^\pm$. Using the fact that $Q^{(m)}(\underline{\delta}_\alpha)_{st} = Q^{(m)}(\underline{\delta}_\alpha)_{s't}$ for all $t \in S$, $m = 1,\ldots,d$ as well as the equality $a = a^+ - a^-$ and the fact that $|r^i(t;\underline{a})| \leq M$ for all $t \in S$, $\underline{a} \in C(t)$ and $i = 1,\ldots,N$, we obtain:

$$|V_\alpha^i(\underline{\delta}^{(\infty)};s) - V_\alpha^i(\underline{\delta}^{(\infty)};s')| \leq \Big|\sum_{m=1}^{\lambda}\sum_{n=0}^{\infty}\alpha^n \sum_{t\in S} r^i(t;\underline{\delta}_\alpha(t)).$$

$$.\{P^{nd+m}(\underline{\delta}_\alpha)_{st} - P^{nd+m}(\underline{\delta}_\alpha)_{s't}\}\Big| \leq$$

$$\leq \Big|\sum_{m=1}^{\lambda}\sum_{n=0}^{\infty}\alpha^n \sum_{t\in S} r^i(t;\underline{\delta}_\alpha(t)).\{P^{nd+m}(\underline{\delta}_\alpha)_{st} - Q^{(m)}(\underline{\delta}_\alpha)_{st}\}^+\Big| +$$

$$+ \Big|\sum_{m=1}^{\lambda}\sum_{n=0}^{\infty}\alpha^n \sum_{t\in S} r^i(t;\underline{\delta}_\alpha(t)).\{P^{nd+m}(\underline{\delta}_\alpha)_{st} - Q^{(m)}(\underline{\delta}_\alpha)_{st}\}^-\Big| +$$

$$+ \Big| \sum_{m=1}^{\lambda} \sum_{n=0}^{\infty} \alpha^n \sum_{t\in S} r^i(t;\underline{\delta}_\alpha(t)).\{P^{nd+m}(\underline{\delta}_\alpha)_{s't} - Q^{(m)}(\underline{\delta}_\alpha)_{s't}\}^+ \Big|$$

$$+ \Big| \sum_{m=1}^{\lambda} \sum_{n=0}^{\infty} \alpha^n \sum_{t\in S} r^i(t;\underline{\delta}_\alpha(t)).\{P^{nd+m}(\underline{\delta}_\alpha)_{s't} - Q^{(m)}(\underline{\delta}_\alpha)_{s't}\}^- \Big| \leq 4\frac{M\nu d}{\rho}$$

where the last inequality follows from (3.6).

To prove (b), fix $s \in C^\beta(\underline{\delta}_\alpha)$ and $s' \in C^\gamma(\underline{\delta}_\alpha)$ such that $\gamma - \beta = m$ (modulo λ). Note that in view of (3.2) and using part (a):

$$|V_\alpha^i(\underline{\delta}_\alpha^{(\infty)};s) - V_\alpha^i(\underline{\delta}_\alpha^{(\infty)};s')| \leq \sum_{\ell=1}^{m} \alpha^\ell \sum_{t\in S} |r^i(t;\underline{\delta}_\alpha(t))|\, P^\ell(\underline{\delta}_\alpha)_{st} +$$

$$+ \alpha^m \sum_{t\in C^\gamma(\underline{\delta}_\alpha)} |V_\alpha^i(\underline{\delta}_\alpha^{(\infty)};t) - V_\alpha^i(\underline{\delta}_\alpha^{(\infty)};s')|\, P^m(\underline{\delta}_\alpha)_{st} +$$

$$+ (1-\alpha^m)\, |V_\alpha^i(\underline{\delta}_\alpha^{(\infty)};s')| \leq$$

$$\leq mM + 4M\nu d/\rho + (1+\alpha+\ldots+\alpha^{m-1})\, M \leq (4\nu+2\rho)\, Md/\rho.$$

Next, let τ be the Markov time, defined by $\tau = \inf\{n \mid s_n \in R(\underline{\delta}_\alpha)\}$ where $\{s_n\}_{n=1}^{\infty}$ denotes the Markov chain associated with the policy $\underline{\delta}_\alpha^{(\infty)}$. Observe from lemma 3.1, part (d) that $E\tau \leq \nu d/\rho$, such that using (b) we obtain for all $s \in S$, and $s' \in R(\underline{\delta}_\alpha)$:

$$|V_\alpha^i(\underline{\delta}_\alpha^{(\infty)};s) - V_\alpha^i(\underline{\delta}^{(\infty)};s)| \leq E_\tau\{\sum_{\ell=1}^{\tau} \alpha^\ell \sum_{t\in S} |r^i(t;\underline{\delta}_\alpha(t))|\, P^\ell(\underline{\delta}_\alpha)_{st} +$$

$$+ \alpha^\tau \sum_{t\in R(\underline{\delta}_\alpha)} |V_\alpha^i(\underline{\delta}^{(\infty)};t) - V_\alpha^i(\underline{\delta}_\alpha^{(\infty)};s')|\, P^\tau(\underline{\delta}_\alpha)_{st} +$$

$$+ (1-\alpha^\tau)\, |V_\alpha^i(\underline{\delta}_\alpha^{(\infty)};s')| \leq$$

$$\leq E\tau.M + (4\nu+2\rho)\, Md/\rho + M\, E_\tau\{1+\alpha+\ldots+\alpha^{\tau-1}\} \leq$$

$$\leq (6\nu+2\rho)\, Md/\rho$$

which proves (c). □

We now prove the existence of an AEP, making use of a technique introduced by TAYLOR [25], and used inter alia in ROSS [17].

THEOREM 2. *Suppose that* A1 *or* A2 *holds. Then there exists a stationary* AEP $\underline{\delta}^{(\infty)}_{-i} \in \underline{\Delta}$, *and for each player* $i = 1,\ldots,N$ *a constant* g^i *and a bounded function* $v^i(\cdot)$ *such that*

$$(3.13)\qquad g^i + v^i(s) = \max_{\mu \in M(A^i(s))} \{r^i(s;[\underline{\delta}^{-i}(s),\mu]) + \sum_{t=1}^{\infty} q_{st}([\underline{\delta}^{-i}(s),\mu])v^i(t)\},$$

for all $s \in S$

where $\delta^i(s)$ *attains the maximum on the right-hand side of* (3.13) *for all* $s \in S$. *Moreover,* $g^i(\underline{\delta}^{(\infty)};s) = g^i$, *for all* $s \in S$, $i = 1,\ldots,N$.

PROOF. We first observe that $|(1-\alpha)\, V^i_\alpha(\underline{\delta}^{(\infty)}_{-\alpha};s^*)| \leq M$ for all $\alpha \in [0,1)$ and $i = 1,\ldots,N$. This together with lemma 3.2 and the fact that for all $s \in S$, any sequence of points in the compact metric space $M(C(s))$ has a convergent subsequence, imply, using the diagonalization procedure, the existence of N constants g^i, N bounded functions $v^i(\cdot)$, a policy $\underline{\delta}^{(\infty)} \in \underline{\Delta}$ and a sequence $\{\alpha_k\}_{k=1}^{\infty}$, with $\alpha_k \in [0,1)$ and $\lim_{k\to\infty} \alpha_k = 1$, such that:

(a) $\lim_{k\to\infty} \underline{\delta}_{\alpha_k} = \underline{\delta}$.

(b) $\lim_{k\to\infty} (1-\alpha_k)\, V^i_{\alpha_k}(\underline{\delta}^{(\infty)}_{-\alpha_k};s^*) = g^i$, $i = 1,\ldots,N$.

(c) $\lim_{k\to\infty} v^i_{\alpha_k}(s) = v^i(s)$, for all $s \in S$, $i = 1,\ldots,N$.

Now, fix $i \in \{1,\ldots,N\}$, and $s = s_0 \in S$ and subtract $V^i_{\alpha_k}(s^*)$ from both sides of (2.7) with $\alpha = \alpha_k$, and $s = s_0$, in order to obtain (cf. (3.11)):

$$(3.14)\qquad v^i_{\alpha_k}(s_0) = \max_{\mu \in M(A^i(s_0))} \{r^i(s_0;[\underline{\delta}^{-i}_{-\alpha_k}(s_0),\mu]) - (1-\alpha_k)\, V^i_{\alpha_k}(s^*) +$$

$$+ \sum_{t=1}^{\infty} q_{s_0 t}([\underline{\delta}^{-i}_{-\alpha_k}(s_0),\mu])\, v^i_{\alpha_k}(t)\}$$

where $\delta^i_{\alpha_k}(s_0)$ attains the maximum on the right-hand side of (3.14). Letting k tend to infinity in (3.14) we obtain (3.13) with $\delta^i(s_0)$ attaining the maximum on the right-hand side of (3.13), as a consequence of (a), (b) and

(c), (2.1) and proposition 18 on p.232 in ROYDEN [18].

Next, it follows from th. 6.17 in ROSS [17] that policy $\underline{\delta}^{(\infty)}$ is an AEP and that $g^i(\underline{\delta}^{(\infty)};s) = g^i$ for all $s \in S$ and $i = 1,\ldots,N$. □

The proof of theorem 2 also shows the following corollary:

COROLLARY 3.3. *If either* A1 *or* A2 *is satisfied, then each limit policy obtained from a sequence of* α-DEPs *with discount factor tending to one, is an* AEP. □

We conclude this section by observing that the existence of an AEP was recently proven under assumption A1.3 in STERN [24].
We note in addition, that condition A1.3 can be weakened as follows:

A1.3': For each policy $\underline{\delta}^{(\infty)} \in \underline{\Delta}$ there exists a state $\underline{s}_\delta$, such that the mean first passage time $\underline{m}_\delta(s,s_\delta)$ is finite and uniformly bounded in $s \in S$, and $\underline{\delta}^{(\infty)} \in \underline{\Delta}$.

The fact that under A1.3' the family of functions $\{v_\alpha^i(\cdot) \mid 0 \leq \alpha < 1\}$ is uniformly bounded, follows from the proof of th. 6.29 in ROSS [17], such that theorem 2 and hence the existence of an AEP, applies to this condition as well.

4. STOCHASTIC GAMES WITH A FINITE STATE AND ACTION SPACE

In this section, we finally consider the N-person stochastic games with finite state and action space, as studied in ROGERS [6] and SOBEL [21].

We first need the following supplementary notations:
Let $A^i(s) = \{1,\ldots,K^i(s)\}$ and let δ^i_{sk}, for any policy $\underline{\delta} \in \underline{\Delta}$, denote the probability with which the kth alternative ($1 \leq k \leq K^i(s)$) is chosen by player i when entering state $s \in S$.

For any policy $\underline{\delta} \in \underline{\Delta}$, we define the fundamental matrix $Z(\underline{\delta}) = [I-P(\underline{\delta})+P^*(\underline{\delta})]^{-1}$ and for each $i = 1,\ldots,N$ the bias-vector $w^i(\underline{\delta})$ by (cf. BLACKWELL [3]):

$$w^i(\underline{\delta})_s = \sum_t Z(\underline{\delta})_{st}[r^i(t;\underline{\delta}(t)) - g^i(\underline{\delta}^{(\infty)};t)].$$

Observe that for each $\underline{\delta} \in \underline{\Delta}$, $g^i(\underline{\delta}^{(\infty)};s) = \sum_t P^*(\underline{\delta})_{st}\, r^i(t;\underline{\delta}(t))$ for

all $i = 1,\ldots,N$, $s \in S$, and that: (cf. [3])

$$(4.1) \qquad V_\alpha^i(\underline{\delta}^{(\infty)};s) = \frac{g^i(\underline{\delta}^{(\infty)};s)}{1-\alpha} + w^i(\underline{\delta})_s + o^i(\alpha;\underline{\delta})_s, \text{ for all } i = 1,\ldots,N,$$

$$s \in S,\ \alpha \in [0,1),$$

where $|o^i(\alpha;\underline{\delta})_s|$ decreases monotonically to zero as $\alpha \uparrow 1$.

Denote by $n(\underline{\delta})$ the number of subchains (closed, irreducible sets of states) for $P(\underline{\delta})$ and let $C^m(\underline{\delta})$ indicate the mth subchain ($1 \le m \le n(\underline{\delta})$). Finally, let $\underline{\Delta}_p \subseteq \underline{\Delta}$ denote the *finite* set of pure and stationary policies and define (cf. SCHWEITZER & FEDERGRUEN [20]):

$$(4.2) \qquad R^* = \{s \mid s \in R(\underline{\delta}) \text{ for some policy } \underline{\delta} \in \underline{\Delta}_p\},$$

the set of states that are recurrent under some pure policy.

Although the existence of an α-DEP is always guaranteed, it is known from a well-known counterexample by GILLETTE [10] that even in the two person-zero sum case an AEP does not need to exist when for some of the policies $\underline{\delta}^{(\infty)} \in \underline{\Delta}$, $P(\underline{\delta})$ is multichained (i.e. $n(\underline{\delta}) \ge 2$). This seeming contrast with the Markov Decision Processes (MDPs) with finite state and action space is explained by the fact that in stochastic games, as distinct from the former, an essential use is made of the set of all randomized actions, whereas in addition the above result perfectly corresponds with what is known to be the case in MDPs with a finite state space, but arbitrary compact action space (cf. BATHER [2]). Under the assumption that for each $\underline{\delta}^{(\infty)} \in \underline{\Delta}_p$, $P(\underline{\delta})$ is unichained, the existence of an AEP was first proved in ROGERS [16] and SOBEL [21]. Moreover, in SOBEL [22], as a still stronger property, the existence of a (g,w)- or bias-equilibrium policy $\underline{\delta}^* \in \underline{\Delta}$ was treated, which we believe should be defined as an AEP $\underline{\delta}^*$, for which:

$$(4.3) \qquad w^i(\underline{\delta}^*)_s \ge w^i(\underline{\eta})_s \text{ for all } i = 1,\ldots,N,\ s \in S \text{ and } \underline{\eta} \in \Pi^{-i}(\underline{\delta}^*) \cap \Pi_{AEP}(\underline{\delta}^*),$$

where

$$\Pi_{AEP}(\underline{\delta}^*) = \{\underline{\eta} \in \underline{\Pi} \mid g^i(\underline{\eta})_s = g^i(\underline{\delta})_s \text{ for all } s \in S\}$$

(the definition 3 in [21] does not extend the (g,w)-optimality notion in Markov Decision Theory; moreover, with the definition in [22], a (g,w)-optimal policy does not even need to exist in the case N = 1, i.e. in the case of an MDP).

In SOBEL [22], the question of the existence of a (g,w)-equilibrium policy was treated using the Brouwer fixed-point theorem with respect to the point-to-point mapping Φ: $\underline{\Delta} \to \underline{\Delta}$, with for all $i \in \psi$, $s \in S$ and $k \in A$:

$$\Phi(\underline{\delta})^i_{sk} = (\delta^i_{sk} + \phi^i_{sk}(\underline{\delta}))/(1 + \sum_{\ell \in A} \phi^i_{s\ell}(\underline{\delta})),$$

where

$$\phi^i_{sk}(\underline{\delta}) = a^i_{sk} + b^i_{sk} + c^i_{sk},$$

and

$$(1) \quad a^i_{sk} = \max\{0, \sum_{t \in S} q_{st}([\delta^{-i}(s),k])\, g^i(\underline{\delta}^{(\infty)};t) - g^i(\underline{\delta}^{(\infty)};s)\},$$

$$(2) \quad b^i_{sk} = \begin{cases} 0, & \text{if } \sum_s \sum_k a^i_{sk} > 0, \\ \max\{0, r^i(s;[\delta^{-i}(s),k]) + \sum_t q_{st}([\delta^{-i}(s),k])\, w^i(\underline{\delta})_t - \\ \qquad - g^i(\underline{\delta}^{(\infty)};s) - w^i(\underline{\delta})_s\}, & \text{otherwise.} \end{cases}$$

$$(3) \quad c^i_{sk} = \begin{cases} 0, & \text{if } \sum_s \sum_k a^i_{sk} > 0, \\ \max\{0, \sum_t q_{st}([\delta^{-i}(s),k])\, z^i(\underline{\delta})_t - w^i(\underline{\delta})_s - z^i(\underline{\delta})_s, & \text{otherwise.} \end{cases}$$

where $z^i(\underline{\delta}) = -Z(\underline{\delta})\, w^i(\underline{\delta})$.

Unfortunately, the mapping Φ may be discontinuous in $\underline{\delta}$, since the $\phi^i_{sk}(\underline{\delta})$ can be discontinuous in those $\underline{\delta}$ that satisfy, for all $i = 1,\ldots,N$, $s \in S$ the functional equation:

$$(4.4) \qquad g^i(\underline{\delta}^{(\infty)};s) = \max_{k \in A^i(s)} \sum_t q_{st}([\delta^{-i}(s),k])\, g^i(\underline{\delta}^{(\infty)};t),$$

or the functional equation (4.5)

$$(4.5)\qquad w^i(\underline{\delta})_s + g^i(\underline{\delta}^{(\infty)};s) = \max_{k\in A^i(s)} \{r^i(s;[\delta^{-i}(s),k]) + \sum_t q_{st}([\delta^{-i}(s),k])w^i(\underline{\delta})_t\},$$

but for which, in any sphere in $\underline{\Delta}$ containing $\underline{\delta}$, policies $\underline{\eta}$ can be found that do *not* satisfy (4.4) (or (4.5) respectively). (As an example consider the MDP with S = {1,2,3}, A = {1,2,3}, $q_{11}(\cdot) = q_{22}(\cdot) = 1$; $q_{31}(1) = q_{31}(2) = 1$; $q_{32}(3) = 1$; $r(1,\cdot) = 1$; $r(2,\cdot) = 0$; $r(3,1) = -M$; $r(3,2) = r(3,3) = 0$; where $M >> 0$. Let δ_x denote the policy that chooses action 1 in state 1 and 2 with probability one, and in state 3 with probability x, whereas in state 3 action 3 is chosen with probability 1-x. Observe that $\phi^1_{32}(\delta)$ is discontinuous in δ_1.)

While under the assumption in SOBEL [22] that $P(\underline{\delta})$ is unichained for every policy $\underline{\delta} \in \underline{\Delta}_p$, the proof in [22] can be rectified in order to show the existence of an AEP (merely by rectifying $\Phi^i_{sk}(\underline{\delta}) = b^i_{sk}$ since in this case only criterion (2) is needed), we observe that this result follows immediately from theorem 2 and the observation that with S a finite state space, the simultaneous Doeblin condition, and hence assumption A1.1 is automatically satisfied.

We note that in both the counterexamples (to the existence of an AEP) by BATHER [2], example 2.3 and GILLETTE [10], the matrix $P^*(\underline{\delta})$ is discontinuous in $\underline{\delta} \in \underline{\Delta}$.

In this section we show in fact that the existence of an AEP is guaranteed, if either $P^*(\underline{\delta})$ is a continuous (matrix)-function on $\underline{\Delta}$, or if the Markov Decision Process that results for any player $i \in \{1,\ldots,N\}$ when the other players have chosen some stationary policy, is a *communicating* system (cf. BATHER [2] and condition B.2 below). Moreover we show that the former property is met under condition B.1 below which is an assumption upon the chain structure of the pure (stationary) policies.

In addition, the approach used in this section has again the advantage of showing that AEPs appear as limit policies from a sequence of α-DEPs with discount factor α tending to one.

Let $\underline{\delta}_1,\ldots,\underline{\delta}_L$ be an enumeration of $\underline{\Delta}_p$, and consider the following

equivalence relation on (cf. SCHWEITZER & FEDERGRUEN [20], proof of Th.3.2):

$$C = \{C^m(\underline{\delta}^r) \mid 1 \le r \le L;\ 1 \le m \le n(\underline{\delta}^r)\}.$$

Let $C \simeq C'$ if there exists $\{C^{(1)} = C, C^{(2)}, \ldots, C^{(n)} = C'\}$ with $C^{(i)} \in C$, and $C^{(i)} \cap C^{(i+1)} \neq \emptyset$, for $i = 1,\ldots,n-1$.

Let $C^{(1)},\ldots,C^{(n^*)}$ be the corresponding equivalence classes on C, and let $R^{*(1)},\ldots,R^{*n^*}$ be the corresponding partition of R^* (cf. (4.2)):

$$R^{*(\ell)} = \cup_{\{(m,r)\,|\,C^m(\underline{\delta}^r)\in C^{(\ell)}\}} C^m(\underline{\delta}^r).$$

The following lemma shows that under assumption B.1, all policies in $\underline{\Delta}$ have the same number of subchains, i.e. $n(\underline{\delta})$ is constant on $\underline{\Delta}$:

B.1. Every (pure) policy $\underline{\delta} \in \underline{\Delta}_p$ has exactly one subchain within each $R^{*(\ell)}$, $\ell = 1,\ldots,n^*$.

LEMMA 4.1. *If* B.1 *holds, then all the policies in* $\underline{\Delta}$ *have the same number of subchains.*

PROOF. Fix $\underline{\delta}^0 \in \underline{\Delta}$. We prove that $P(\underline{\delta}^0)$ has exactly one subchain within each $R^{*(\ell)}$ (ℓ=1,...,n) by showing subsequently:
(1) $R(\underline{\delta}^0) \subseteq R^*$;
(2) any subchain of $P(\underline{\delta}^0)$ is contained within one of the sets $R^{*(\ell)}$;
(3) in every one of the sets $R^{*(\ell)}$ there is exactly one subchain of $P(\underline{\delta}^0)$.
(1) and (2) follow immediately from parts (a) and (c) of Th. 3.2 in [19], so that (3) remains to be shown.

Fix ℓ ($1 \le \ell \le \eta^*$) and assume first that $R(\underline{\delta}^0) \cap R^{*(\ell)} = \emptyset$. It then follows from Lemma 2.2 in [20] that there exists a pure policy $\underline{\eta} \in \underline{\Delta}_p$, with $R(\underline{\eta}) \subseteq R(\underline{\delta}^0)$, such that $R(\underline{\eta}) \cap R^{*(\ell)} = \emptyset$, contradicting B.1. Finally, observe that for any pair $\underline{\delta}_1, \underline{\delta}_2 \in \underline{\Delta}_p$, the subchains of $\underline{\delta}_1$ and $\underline{\delta}_2$ that are contained within $R^{*(\ell)}$ must intersect, since it would otherwise be possible to construct a $\underline{\delta}_3 \in \underline{\Delta}_p$ with two subchains within $R^{*(\ell)}$, contradicting B.1, and verify that this property implies that $P(\underline{\delta})$ cannot have two or more subchains within $R^{*(\ell)}$.

REMARK. Assume that every policy in $\underline{\Delta}_p$ is unichained (cf. SOBEL [22], ROGERS [16]) and observe that this assumption implies for any pair $(\underline{\delta}_1, \underline{\delta}_2) \in \underline{\Delta}_p$ that their subchains must intersect, so that all the subchains in C belong to the same equivalence class, i.e. $n^* = 1$.

It hence follows that the assumption in SOBEL [22] and ROGERS [16] is identical with the special case of B.1 where $n^* = 1$.

We next introduce assumption B.2:

B.2. For every $i \in \{1,\ldots,N\}$; for every pair of states $s,t \in S$, and for every combination $\{\delta^j \in \Delta \mid j \neq i\}$ of the other players, there is a policy $\delta^i \in \Delta$ for player i and an integer ℓ such that $P(\underline{\delta})^{\ell}_{st} > 0$, which can be seen as an extension of the *communicatingness*-property (cf. BATHER [2], HORDIJK [12]).

It is easily verified (cf. BATHER [2], part II, p.526) that under assumption B.2 the seemingly stronger condition (4.6) is satisfied.

(4.6) for every $i \in \{1,\ldots,N\}$ and for every combination $\{\delta^j \in \Delta \mid j \neq i\}$ of the other players there is a policy $\delta^i \in \Delta$ for player i, such that $P(\underline{\delta})$ is an irreducible Markov Chain, where $\underline{\delta} = [\delta^1,\ldots,\delta^N]$.

Using the fact that in an irreducible Markov Chain the mean first passage time from any state s to any state t is finite one concludes that B.2 is in fact the relaxation of assumption A.2 to the finite state space model.

Theorem 3 below proves, under B.1 as well as under B.2 the existence of an AEP.

THEOREM 3. *There exists a stationary* AEP, *if either* B.1 *or* B.2 *holds.*

PROOF. Assume first that B.1 holds. Fix $i = 1,\ldots,N$, $s \in S$. It follows from Lemma 4.1 that $n(\underline{\delta})$ is constant on $\underline{\Delta}$, and hence from Th. 5 in SCHWEITZER [19] that $P^*(\underline{\delta})$ is continuous in $\underline{\delta} \in \underline{\Delta}$, which in its turn invokes, by their very definition, the continuity of $g^i(\underline{\delta}^{(\infty)};s)$ and $w^i(\underline{\delta})_s$ in $\underline{\delta} \in \underline{\Delta}$.

We first fix an α-DEP $\underline{\delta}_\alpha \in \underline{\Delta}$, for each $\alpha \in [0,1)$.

Inserting (4.1) into both sides of (1.8) and multiplying both sides of the resulting inequality by $(1-\alpha)$ we obtain for all $\underline{\eta} \in \underline{\Delta}$

$$(4.7) \qquad g^i(\underline{\delta}_\alpha^{(\infty)};s) + (1-\alpha)\, w^i(\underline{\delta}_\alpha)_s + (1-\alpha)\, o^i(\alpha;\underline{\delta}_\alpha)_s \geq$$

$$\geq g^i([\delta_\alpha^{-i},\eta]^{(\infty)};s) + (1-\alpha)\, w^i([\delta_\alpha^{-i},\eta])_s + (1-\alpha)\, o^i(\alpha;[\delta_\alpha^{-i},\eta])_s .$$

It next follows from the fact that $\underline{\Delta}$ is a compact metric space that one can find a policy $\underline{\delta}^{*(\infty)} \in \underline{\Delta}$, and a sequence $\{\alpha_k\}_{k=1}^{\infty}$, with $\alpha_k \in [0,1)$ and $\lim_{k\to\infty} \alpha_k = 1$, such that $\lim_{k\to\infty} \underline{\delta}_{\alpha_k} = \underline{\delta}^*$. We further show:

$$(4.8) \qquad \lim_{k\to\infty} (1-\alpha_k)\, o^i(\alpha_k;\underline{\delta}_{\alpha_k})_s = 0 = \lim_{k\to\infty} (1-\alpha_k)\, o^i(\alpha_k;[\delta_{\alpha_k}^{-i},\eta])_s .$$

Merely proving the first equality in (4.8) (the proof of the second one being analogous), we observe that for each $\alpha \in [0,1)$, $o^i(\alpha;\underline{\delta})_s$ is continuous in $\underline{\delta} \in \underline{\Delta}$, as a result of Lemma (2.2), relation (4.1) and the continuity of $g^i(\underline{\delta}^{(\infty)};s)$ and $w^i(\underline{\delta})_s$ in $\underline{\delta} \in \underline{\Delta}$.

(4.8) then follows from the fact that for any $\underline{\eta} \in \underline{\Delta}$, $|(1-\alpha)\, o^i(\alpha;\underline{\eta})_s|$ decreases monotonically to zero, as $\alpha \to 1$, using e.g. Dini's Theorem (cf. ROYDEN [18], p.162).

Finally, let k tend to infinity on both sides of (4.7) with $\alpha = \alpha_k$, and use (4.8) as well as the continuity of $g^i(\underline{\delta}^{(\infty)};s)$ and $w^i(\underline{\delta})_s$ in $\underline{\delta} \in \underline{\Delta}$, in order to obtain:

$$(4.9) \qquad g^i(\underline{\delta}^{*(\infty)};s) \geq g^i([\delta^{*-i},\eta]^{(\infty)};s), \qquad \text{for all } i = 1,\ldots,N;$$

$$s \in S \text{ and } \eta \in \Delta^i .$$

Consider next the "decision problem" that arises when all players but player i tie themselves down to their respective policies in $\underline{\delta}^*$, and observe from (4.8) that in this decision problem, δ^{*i} is a maximal gain policy to player i within Δ. It then follows from Theorem 2 in BLACKWELL [4] that δ^{*i} is also optimal within Π^i. This proves the theorem under B.1, whereas the existence of an AEP under B.2 follows immediately from Theorem 2, B.2 being the relaxation of A.2 to the finite space model. □

We finally turn to the question under which condition(s) a pure instead of a randomized AEP exists, for every choice of the one-step expected rewards $r^i(s;\underline{a})$.

So far the only stochastic games known to have this property are the so-called two person-zero sum games with perfect information, in which in each state of the system one of the two players has not more than one alternative.

The existence of a pure AEP for this class of stochastic games was first treated by GILLETTE [10]. Unfortunately an incorrect extension of the Hardy-Littlewood theorem was used, as has been pointed out by LIGGETT & LIPPMAN [15].

The existence of a pure AEP, and, as an even stronger result, the existence of a pure bias-equilibrium policy may, however be derived from the fact that a pure stationary α-DEP exists for each $\alpha \in [0,1)$, where the latter has already been proved by SHAPLEY [21].

Since $\underline{\Delta}_p$ is a finite set, we can therefore find a policy $\underline{\delta}^* = (\delta^{*1},\delta^{*2}) \in$ $\in \underline{\Delta}_p$ and a sequence $\{\alpha_n\}_{n=1}^{\infty}$, with $\alpha_n \uparrow 1$, such that $\underline{\delta}^*$ is an α_n-DEP for $n = 1,2,\ldots$. Let $r(s;\underline{a}) = r^1(s;\underline{a}) = -r^2(s;\underline{a})$ and $V_\alpha(\underline{\eta};s) = V_\alpha^1(\underline{\eta};s) =$ $= -V_\alpha^2(\underline{\eta};s)$, and observe that $V_\alpha(\underline{\eta};s) = \sum_t [I-\alpha P(\underline{\eta})]^{-1}_{st} r(t;\underline{\eta}(t))$ is a rational function in α for all $\underline{\eta} \in \underline{\Delta}_p$ and $s \in S$.

Since $V_\alpha([\eta^1,\delta^{*2}];s) - V_\alpha(\underline{\delta}^*;s)$ and $V_\alpha([\delta^{*1},\eta^2];s) - V_\alpha(\underline{\delta}^*;s)$ are also rational functions in α, for all $\eta^1,\eta^2 \in \Delta$ and $s \in S$, and hence are either identically zero or have a finite number of zeros, there exists an $\tilde{\alpha}(\eta^1,\eta^2,s)$ such that, for all $\alpha > \tilde{\alpha}(\eta^1,\eta^2,s)$:

$$(4.9) \qquad V_\alpha([\eta^1,\delta^{*2}];s) \leq V_\alpha(\delta^*;s) \leq V_\alpha([\delta^{*1},\eta^2];s).$$

Since S and $\underline{\Delta}_p$ are finite, we thus obtain an α^* such that $\underline{\delta}^*$ is an α-DEP for all $\alpha > \alpha^*$. It then follows by comparing the Laurent series expansion for $V_\alpha(\underline{\delta}^*)$ and $V_\alpha([\eta^1,\delta^{*2}])$ as well as the one of $V_\alpha(\underline{\delta}^*)$ and $V_\alpha([\delta^{*1},\eta^2])$ that $\underline{\delta}^*$ is a bias-equilibrium policy, and more generally an equilibrium policy under all of the sensitive discount optimality criteria (cf. MILLER & VEINOTT [15a]).

REMARK. The proof in LIGGETT & LIPPMAN [15] for the existence of a pure AEP is more complicated than the one above; moreover, it requires an additional argument. More specifically, instead of th.5 in BLACKWELL [4] we need the stronger result that in each Markov Decision Model there exists a discount factor α^* such that any policy that is α-optimal for some $\alpha > \alpha^*$ is α-optimal for all $\alpha > \alpha^*$, which is immediate from the proof of th.5. Relation (5) in [15] should be adapted in this sense.

One might wonder whether the existence of a pure AEP is also guaranteed in the case of two-person, *nonzero*-sum, or even more generally in the case of N-person games with perfect information. The following two-person game is, however a counterexample, which is due to VRIEZE & WANROOIJ [28]. Let $S = \{1,2\}$ and $A^1(1) = A^2(2) = \{1,2\}$ with $A^2(1) = A^1(2) = \{1\}$. Let $r^2(1;(1,1)) = r^1(2;(1,1)) = 1$ and $r^2(1;(2,1)) = r^1(2;(1,2)) = -1$, the other rewards being zero, and let
$q_{11}(1,1) = q_{21}(1,1) = 2/3$ and $q_{11}(2,1) = q_{21}(1,2) = 1/3$.

Acknowledgement

I am infinitely grateful to both Koos Vrieze and Gerard Wanrooij for many fruitful discussions.

In addition I am indebted to Henk Tijms for his guidance, and numerous helpful comments and suggestions, as well as to Arie Hordijk for stimulating discussions.

REFERENCES

[1] ANTHONISSE, J. & H. Tijms, *On the stability of products of stochastic matrices*, to appear in J. Math. Anal. Appl.

[2] BATHER, J., *Optimal decision procedures for finite Markov chains* I, II, Adv. in Applied Prob.5 (1973).

[3] BILLINGSLEY, P., *Convergence of probability measures*, Wiley, New York (1968).

[4] BLACKWELL, D., *Discrete dynamic programming*, Ann. Math. Stat.36 (1962), 719-726.

[5] BLACKWELL, D., *Discounted dynamic programming*, Ann. Math. Stat.39 (1968), 226-235.

[6] CHUNG, K., *Markov chains with stationary transition probabilities*, second edition, Springer, Berlin (1960).

[7] DERMAN, C. & R. STRAUCH, *A note on memoryless rules for controlling sequential control processes*, Ann. Math. Stat.37 (1966), 276-278.

[8] DOOB, J., *Stochastic Processes*, Wiley, New York (1953).

[9] FAN, K., *Fixed-point and minimax theorems in locally convex topological linear spaces*, Proc. Nat. Acad. Sci.38 (1952), 121-126.

[10] GILLETTE, D., *Stochastic games with zero stop probabilities*, in M. Dresher et al. (eds.), *Contributions to the theory of games*, Vol. III (Princeton Univ. Press, Princeton, New Jersey (1957), 179-188.

[11] GLICKSBERG, I., *A further generalization of the Kakutani fixed point theorem with application to Nash equilibrium points*, Proc. Amer. Math. Soc.3 (1952), 170-174.

[12] HORDIJK, A., *Dynamic Programming and Markov Potential Theory*, MC Tract 51, Mathematisch Centrum, Amsterdam (1974).

[13] HORDIJK, A., O. VRIEZE & G. WANROOIJ, *Semi-Markov strategies in stochastic games* (1976), to appear.

[14] IDZIK, A., *Two-stage noncooperative discounted stochastic games*, Computation Centre of the Polish Academy of Sciences, Warsaw (1976).

[15] LIGGETT, T. & S. LIPPMAN, *Stochastic games with perfect information and time average payoff*, SIAM Review 11 (1969), 604-607.

[15a] MILLER, B. & A. VEINOTT Jr., *Discrete Dynamic programming with a small interest rate*, Ann. Math. Stat. 40 (1969), 366-370.

[16] ROGERS, P., *Nonzero-sum stochastic games*, Report ORC 69-8, Operations Res. Center, Univ. of California, Berkeley, Calif. (1969).

[17] ROSS, S., *Applied Probability Models with Optimization Applications*, Holden-Day, San Fransisco, Calif. (1970).

[18] ROYDEN, H., *Real Analysis*, 2nd ed., MacMillan, New York (1968).

[19] SCHWEITZER, P.J., *Perturbation theory and finite Markov chains*, J. Appl. Prob. (1968), 401-413.

[20] SCHWEITZER, P.J. & A. FEDERGRUEN, *The functional equations of undiscounted Markov renewal programming*, (1975), to appear in Math. of Oper. Res.

[21] SHAPLEY, L., *Stochastic games*, Proc. Nat. Acad. Sci. U.S.A. 39 (1953), 1095-1100.

[22] SOBEL, M., *Noncooperative stochastic games*, Ann. of Math. Stat. 42 (1971), 1930-1935.

[23] SOBEL, M., *Continuous stochastic games*, J. Appl. Prob. 10 (1973), 597-604.

[24] STERN, M., *On stochastic games with limiting average payoff*, Ph.D. dissertation, Dept. of Math., Univ. of Illinois, Chicago Circle Campus (1975).

[25] TAYLOR, H., *Markovian sequential replacement processes*, Ann. Math. Stat.36 (1965), 1677-1694.

[26] TIJMS, H., *On dynamic programming with arbitrary state space, compact action space and the average return as criterion*, Report BW 55/75, Mathematisch Centrum, Amsterdam (1975).

[27] VARADARAJAN, V., *Weak convergence of measures on separable metric spaces*, Sankhya (1958), 15-22.

[28] VRIEZE, O. & G. WANROOIJ, Private communication, (1975).